MOLECULAR ECOLOGY

Molecular Ecology (2016) doi: 10.1111/mec.13887

Spatio-temporal analysis of Nova virus, a divergent hantavirus circulating in the European mole in Belgium

LIES LAENEN,* SIMON DELLICOUR,† VALENTIJN VERGOTE,* INNE NAUWELAERS,* SARAH DE COSTER,* INA VERBEECK,* BERT VANMECHELEN,* PHILIPPE LEMEY† and PIET MAES*

*KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory of Clinical Virology, Herestraat 49, 3000 Leuven, Belgium, †KU Leuven, Department of Microbiology and Immunology, Rega Institute for Medical Research, Laboratory of Evolutionary and Computational Virology, Herestraat 49, 3000 Leuven, Belgium

Abstract

Over the last decade, the recognized host range of hantaviruses has expanded considerably with the discovery of distinct hantaviruses in shrews, moles and bats. Unfortunately, in-depth studies of these viruses have been limited. Here we describe a comprehensive analysis of the spatial distribution, genetic diversity and evolution of Nova virus, a hantavirus that has the European mole as its natural host. Our analysis demonstrated that Nova virus has a high prevalence and widespread distribution in Belgium. While Nova virus displayed relatively high nucleotide diversity in Belgium, amino acid changes were limited. The nucleocapsid protein was subjected to strong purifying selection, reflecting the strict evolutionary constraints placed upon Nova virus by its host. Spatio-temporal analysis using Bayesian evolutionary inference techniques demonstrated that Nova virus had efficiently spread in the European mole population in Belgium, forming two distinct clades, representing east and west of Belgium. The influence of landscape barriers, in the form of the main waterways, on the dispersal velocity of Nova virus was assessed using an analytical framework for comparing Bayesian viral phylogenies with environmental landscape data. We demonstrated that waterways did not act as an environmental resistance factor slowing down Nova virus diffusion in the mole population. With this study, we provide information about the spatial diffusion of Nova virus and contribute sequence information that can be applied in further functional studies.

Keywords: evolution, Nova hantavirus, Talpidae, zoonosis

Received 17 August 2016; revision received 10 October 2016; accepted 12 October 2016

Introduction

Zoonoses form a large portion of emerging infectious diseases threatening global health. Over 70% have their origin in wildlife and are increasing significantly over time (Taylor *et al.* 2001; Jones *et al.* 2008). Special attention is given to viral zoonoses and RNA viruses in particular, due to their more error-prone replication and consequential ability to adapt to new hosts more easily (Cleaveland *et al.* 2001; Woolhouse & Gowtage-Sequeria 2005). Hantaviruses are negative sense single-stranded RNA viruses and the causative agents of haemorrhagic fever with renal syndrome (HFRS) and hantavirus pulmonary syndrome (HPS) (Maes *et al.* 2004; Vaheri *et al.* 2013). In Europe and Asia, human infection with Old World hantaviruses can lead to HFRS, which has the kidneys as its main target organs. Cases of HPS, present in the Americas, are rare, albeit they usually have a more severe disease course. Although the organ tropism can differ in both clinical presentations, an apparent overlap of symptoms is occasionally noted (Peters & Khan 2002; Gizzi *et al.* 2013). Hantavirus disease, a more suitable name for all clinical manifestations associated with hantavirus infection, can be contracted upon

Correspondence: Lies Laenen, Piet Maes, Fax: +3216330026, E-mails: lies.laenen@kuleuven.be (LL), piet.maes@kuleuven.be (PM)

inhalation of aerosols of saliva, urine or faeces of infected reservoirs (Clement *et al.* 2012).

At present, all confirmed pathogenic hantaviruses have rodents as their natural hosts. Even though the first hantavirus, Thottapalayam virus, was isolated from an Asian house shrew (*Suncus murinus*) in 1964 (Carey *et al.* 1971), all subsequently discovered hantaviruses until 2006 originated from rodents. Since then, numerous new hantaviruses have been detected in shrews, moles and bats (Klempa *et al.* 2007; Arai *et al.* 2008; Sumibcay *et al.* 2012; Weiss *et al.* 2012). Currently, 24 hantavirus species have been officially recognized, but more than 60 possible hantaviruses have been detected up till now (Plyusnin *et al.* 2011). The discovery of new hantaviruses gave more insight into the complex evolution that led to the hantaviruses circulating today. The congruence between phylogenetic trees of hantaviruses and their rodent hosts inspired an initial hypothesis of codivergence, where hantaviruses presumably had evolved with their specific hosts over millions of years (Hughes & Friedman 2000). The discovery of new hantaviruses of Eulipotyphla and Chiroptera revealed inconsistencies with strict long-term virus-host codivergence (Guo *et al.* 2013). Hantavirus evolution appears to consist of a complex mechanism of virus-host co-evolution that has involved ancient reassortment and multiple host-switching events (Bennett *et al.* 2014). Moreover, ancestors of bats or insectivores instead of rodents are the likely hosts of primordial hantaviruses (Guo *et al.* 2013; Bennett *et al.* 2014). Hantaviruses with Talpidae as host especially have a complex evolutionary history as the mole-borne viruses do not form a monophyletic group and are the result of multiple separate host-switching events (Bennett *et al.* 2014). Hantaviruses have been detected in moles and shrew moles in America, Europe and Asia, for example Asama virus (*Urotrichus talpoides*), Oxbow virus (*Neurotrichus gibbsii*), Rockport virus (*Scalopus aquaticus*), Dahonggou Creek virus (*Scaptonyx fusicaudus*) and Nova virus (*Talpa europaea*) (Arai *et al.* 2008; Kang *et al.* 2009a,b, 2011, 2016). Unfortunately, for most newfound hantaviruses, isolates and/or complete genome sequences are lacking. This causes issues with current hantavirus classification guidelines and hampers infection and pathogenicity studies with these agents.

Nova virus, abbreviated as NVAV, was detected in 2009 in the archival liver tissue of a European mole, captured in Zala County, Hungary (Kang *et al.* 2009b). The European mole has a widespread distribution across the European continent although it is absent from the southern part of the Mediterranean regions and the north of Scandinavia (Amori *et al.* 2008). In line with these observations, Nova virus has been detected in European moles from France, Poland and Belgium (Gu *et al.* 2014a,b; Peeraer *et al.* 2015). The mean prevalence of Nova virus within the mole populations in these countries is high, going up to 65.5%. Nova virus has a wide tissue distribution in lung, heart, liver, kidney, spleen and intestine, concurring with previous observations of rodent-borne hantaviruses and their reservoirs (Gu *et al.* 2014b). Although insectivore-borne viruses have been detected in their natural hosts for almost 10 years, little is still known about their pathogenic potential, mode of transmission, evolutionary history, genetic diversity and host adaptation potential. Nova virus is one of the most divergent hantaviruses detected to date and is distantly related to bat-borne hantaviruses, likely having a more ancient origin than rodent-borne hantaviruses (Guo *et al.* 2013; Bennett *et al.* 2014). Intraperitoneal infection of infant mice with Nova virus resulted in neurological symptoms and death, after which a high number of Nova virus RNA was detected in brain tissue. Notably, histopathological abnormalities were largely confined to the brain and symptoms do not resemble human hantavirus disease progression (Gu *et al.* 2016). Infection studies in more suitable animal models such as immunosuppressed hamsters will be paramount to uncover the pathogenic potential of Nova virus (Brocato *et al.* 2014).

Here we do an in-depth analysis of the genetic characteristics, molecular epidemiology and evolution of Nova virus. Foremost, we investigate the distribution of Nova virus in the Belgian mole population. In addition, we study the genetic diversity of Nova virus in comparison with other hantaviruses and between Belgian Nova virus strains. Furthermore, we examine the evolutionary pressures placed upon the Nova virus nucleocapsid protein. Finally, we analyse the spatio-temporal diffusion of Nova virus in the Belgian mole population and factors influencing virus dispersal velocity. Due to incomplete and sparse sequence data, comprehensive genetic studies of insectivore-borne hantaviruses have been limited (Schlegel *et al.* 2012; Guo *et al.* 2013; Bennett *et al.* 2014; Lin *et al.* 2014). This study provides more insights into the complex evolutionary history of insectivore-borne hantaviruses and will hopefully form a jumping-off point for more extensive studies concerning other bat-, shrew- and mole-borne hantaviruses.

Materials and methods

Sample collection

European moles were trapped at different locations in the period of 2013–2015. Trapping was conducted in the context of pest control by professional mole catchers using mechanical scissor traps. As European moles are widely persecuted as a pest animal, no permits were required for the described study, which complied with

all relevant regulations. Immediately after trapping, European moles were identified by morphological features and stored for up to 1 week at −20 °C until dissection. Lung, kidney, heart, liver and spleen tissues were aseptically removed and stored in RNA*later*® Stabilization Solution (Ambion) at −20 °C.

RNA extraction, RT–PCR and sequencing

Total RNA was extracted from renal tissue with the RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. A nested RT–PCR was carried out using the OneStep RT–PCR kit (Qiagen) using primers described previously (Klempa *et al.* 2006). After an initial RT step at 50 °C for 30 min, Taq polymerase was activated at 95 °C for 15 min. This was followed by 40 cycles of amplification (30 s at 94 °C, 30 s at 53 °C and 1 min at 72 °C) and a final extension at 72 °C for 10 min. Primers for the amplification of the full S segment are available in Table S1 (Supporting information). Additionally, primer walking was performed to cover the full M and L segments of seven Nova virus sequences. Primers and thermal profiles are available upon request. PCR amplicons were purified using ExoSAP-IT® PCR Product Cleanup (Affymetrix) and sequenced according to the ddNTP chain termination method with the BigDye Terminator v3.1 cycle sequencing kit (Life Technologies) on an Applied Biosystems 3130xl Genetic Analyzer. Sequences were manually inspected using CHROMAS 2.4 (Technelysium), and consensus sequences were derived with SEQMAN 7.0 (DNAstar). All sequences generated in this study were submitted to NCBI GenBank under Accession numbers KX512326 to KX512437.

Nucleotide and protein analysis

Nucleotide sequences were aligned with MAFFT using the L-INS-I option and manually edited in MEGA 7 (Kumar *et al.* 2016). Nucleotide and protein identities were inferred using MEGA 7. Interstrain nucleotide diversity (π) was calculated with DNASP using the DNA polymorphism function with a sliding window of 100 nt and a step size of 25 (Librado & Rozas 2009). N- and O-linked glycosylation sites were predicted using the NETNGLYC 1.0 Server and NETOGLYC 4.0 Server, respectively (Steentoft *et al.* 2013). Possible recombination events were analysed using RDP, GENECONV, Chimaera, MaxChi, 3seq, Bootscan and Siscan, implemented in RDP4 (Martin *et al.* 2015). To estimate the genetic similarity between virus pairs, pairwise evolutionary distances (PED) were calculated on the complete nucleocapsid coding region multiple alignment using the program TREE-PUZZLE v5.2 with the WAG amino acid substitution matrix. Unrooted phylogenetic reconstruction was performed with PHYML v3.0 using default parameters and automatic model selection. A heatplot was generated with R v3.1.1. Selection pressure analysis was performed using the HYPHY package implemented in the DATAMONKEY server (Delport *et al.* 2010). Sites under positive and purifying selection were assessed using SLAC, FEL, IFEL, MEME and FUBAR (Kosakovsky Pond & Frost 2005; Pond *et al.* 2006; Murrell *et al.* 2012, 2013).

Phylogeographic analysis

To estimate geo-referenced phylogenetic trees, Bayesian phylogeographic inference was performed using a continuous diffusion approach implemented in BEAST 1.8.2 (Lemey *et al.* 2010). The nucleotide substitution process was modelled using a codon position partition model [SRD06, (Shapiro *et al.* 2006)]. Because the data set did not contain sufficient temporal signal (divergence accumulation over the sampling time range as assessed using TEMPEST, Rambaut *et al.* 2016), we used a strict molecular clock model and fixed the substitution rate to 10^{-3} substitutions/site/year in line with hantavirus evolutionary rate estimates (Ramsden *et al.* 2008). We used a Bayesian SkyGrid coalescence model (Gill *et al.* 2013) and a relaxed random walk model for the continuous location traits with a Cauchy probability distribution. The Markov chain Monte Carlo analyses were run until adequate effective sample sizes (ESS > 200 for continuous parameters) were obtained. A maximum clade credibility tree was summarized from the posterior tree distribution using TREEANNOTATOR 1.8.2 [from the BEAST package; (Drummond *et al.* 2012)].

Testing the impact of main waterways

We used the analytical framework developed by Dellicour *et al.* (2016b) to investigate the impact of main waterways on the dispersal velocity of the Nova virus in the Belgian mole population. We here summarize the different steps of the workflow, but see also for a more detailed description of a similar workflow. All the different analytical steps were performed with R functions available in the package SERAPHIM (Dellicour *et al.* 2016a).

Step 1: Spatial-temporal information was extracted from 100 trees sampled at regular intervals from the post-burn-in posterior distribution to take into account phylogenetic uncertainty. After this extraction step, each phylogeny branch in each tree was summarized as a vector defined by its start and end location and its start and end dates. Each branch

therefore represents an independent viral lineage dispersal event (Pybus *et al.* 2012).

Step 2: Second, to obtain a first visual overview of the extracted information, we mapped both the MCC tree and the 75% HPD regions based on the 100 trees selected in step 1. Similar to the method developed by Bielejec *et al.* (2011), the 75% HPD regions were computed for successive time layers and then superimposed using a specific colour scale ranging from red (MRCA) to blue (sampling time). Main waterways, that is 'rivers' and 'canals' as defined by and obtained from the OPENSTREETMAP database (URL: openstreetmap.org), were also added to the map.

Step 3: In a third step, each of the vectors obtained in step 1 was assigned a specific 'weight' based on main waterway rasters defined as resistance grids. The main waterway rasters were generated by rasterizing the 'rivers' and 'canals' GIS vectorial objects obtained from the OPENSTREETMAP database and by assigning a resistance value equal to $(1 + k)$ to each cell crossed by a main waterway. We tested three different values for the parameter k: 10, 100 and 1000. As the raster cells that are not crossed by a main waterway were assigned a uniform value of 1, k thus defines the additional resistance when the cell does contain such potential landscape barrier. To compute these weights, we used two different 'path models': (i) a least-cost path model where weight is computed as the sum of the transition values between adjacent cells along the least-cost path (Dijkstra 1959) and (ii) a random walk path model where the weight is a graph-theoretic metric based on circuit theory, which takes into account multiple possible pathways connecting a given pair of locations and the values of the raster cells through which they pass (McRae 2006; Mcrae *et al.* 2008). These two models correspond to two different dispersion behaviours, and testing both of them allows the possibility of considering different assumptions about the dynamics of Nova virus spread.

Step 4: In a fourth step, we estimated the correlation between the dispersal duration and environmental weight now assigned to each phylogeny branch. For each of the 100 sampled trees, we estimated the statistic $D = (R^2_{\text{waterways}} - R^2_{\text{null}})$, where $R^2_{\text{waterways}}$ is the R^2 obtained when branch durations are regressed against weights computed on a main waterways raster and R^2_{null} is the R^2 obtained when branch durations are regressed against weights computed on a 'null' raster, that is a raster with uniform cell values equal to 1. R^2_{null} corresponds to the case where only the spatial distance of each movement event is considered.

Step 5: The last step consisted in testing the statistical significance of D values. For this purpose, we used the randomization procedure that randomizes phylogenetic node positions under the constraint that branch lengths (i.e. branch durations), the tree topology and the root position are unchanged (Dellicour *et al.* 2016b). Such a procedure allows generating a null distribution of D values for each sampled tree. The comparison of the observed D value and this null distribution provided one P-value per input phylogeny, which corresponds to the proportion of randomization replicates that generated D values larger than that generated by the empirical input phylogeny. As 100 sampled trees were used, we then obtained a distribution of 100 P-values and reported the percentage of P-values >0.05.

Results

Nova virus has a widespread distribution in Belgium

To gain insight into the spatial spread of Nova virus in the Belgian mole population, we tested mole kidney samples for the presence of Nova virus RNA. From 2013 to 2015, a total of 479 moles were collected in Belgium. The host species was determined by morphological identification. All samples in our data set originated from European moles (*Talpa europaea*), the only mole species described to be present in this region. Samples had a widespread distribution in Belgium, covering 93 unique postal codes (Fig. 1). The range of the sampled regions was 152 km. Kidney samples were tested for the presence of hantavirus RNA with primers directed at a conserved region of the polymerase gene (Klempa *et al.* 2006). Sequencing confirmed the presence of Nova virus RNA in 255 of 479 samples, leading to a positivity rate of 53.2%. The geographical distribution of positive samples included the entire sampling region. There was no significant fluctuation in positivity rates between subsequent years. Kidney and lung samples of two European water voles (*Arvicola amphibius*) that were trapped in mole burrows were also tested for the presence of hantavirus RNA but were found to be negative. These findings indicate that Nova virus has a high prevalence and is widely dispersed in the mole population in Belgium.

Full genome analysis of Nova virus displays high genetic diversity

To further investigate the full genetic diversity of Nova virus, the complete small (S), medium (M) and large (L) segments of seven Nova virus strains, sampled from

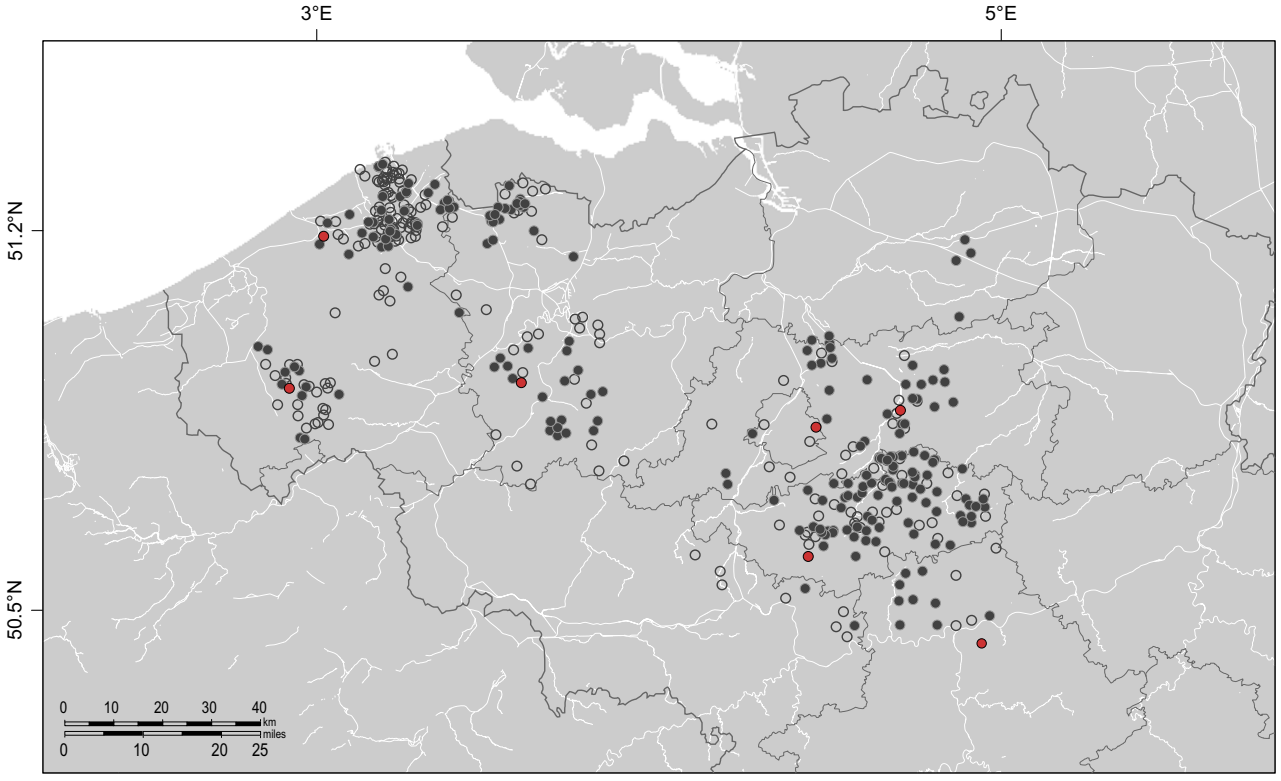

Fig. 1 Map of the distribution of Nova virus in the mole population in Belgium. Nova virus RNA-positive samples were marked with full circles, while negative samples are marked with open circles. Samples with a complete genome sequence are marked in red.

divergent regions, were sequenced (Fig. 1). Within Belgian Nova virus strains, there was relatively high diversity at the nucleotide level, with strain variability going up to 16.25% for the complete S segment, 16.65% for the M segment and 14.50% for the complete L segment (Table 1). The lower nucleotide diversity within strains from outside of Belgium can be explained by the fact that the majority of samples originated from a small region in Poland. However, at the amino acid level, high codon conservation was noticed with maximal amino acid identity variation of 97.89%, 96.36% and 96.34% for S, M and L, respectively. The interstrain nucleotide diversity (π) for the Belgian strains was calculated using a sliding window approach (Fig. 2). The S segment, which had 540 polymorphic sites, demonstrated relative sequence conservation in the coding region and high sequence variability at the 3′ noncoding region (NCR). At the 3′ NCR, two more conserved domains were present around nucleotide 1400 and nucleotide 1650. The M segment, exhibiting 1042 segregating sites, had approximately the same level of overall sequence variability. In addition to an expected peak in nucleotide diversity at the short 3′ NCR, a conserved region was noticed at the 5′ end, which corresponds to the signal peptide and the N-terminal region of the Gn protein. The L segment remains relatively consistent over the entire polymerase gene with 2617 segregating sites. A region with lower variation was present around nucleotide 3400–3500.

The presence of N- and O-linked glycosylation sites was determined for the Nova virus glycoprotein precursor of all seven sequenced samples and one additional sequence available from NCBI GenBank. All glycoproteins had predicted N-glycosylation sites at N101, N133, N344, N396, N539 and N924. The predicted glycosylation at position N101 is a unique characteristic of all Nova viruses and has not been found in other hantaviruses thus far. O-glycosylation was inferred at a number of amino acid positions, but no predicted glycosylation sites were present in all analysed strains (Table S2, Supporting information).

The interspecies divergence between Nova virus and other hantaviruses was characterized by a heatmap based on the pairwise evolutionary distances of the nucleocapsid protein (Fig. 3). A strong grouping of virus pairs in correlation with their natural host was observed. Furthermore, a cluster of hantaviruses with hosts belonging to the Soricidae family (i.e. Kilimanjaro virus, Imjin virus and Thottapalayam virus) proved to be highly divergent of other hantaviruses. A second

Table 1 Table with average nucleotide (nt) and amino acid (aa) identities. The average nt and aa identities were calculated for the Belgian strains sequenced in this study, for non-Belgian strains available on NCBI GenBank (other) and between both groups. The absolute range of nucleotide or amino acid identities is denoted between square brackets. N gives the number of strains included in each analysis

	N	Avg. nt identity	Avg. aa identity
	Nova virus S segment		
Within Belgium	7	85.68% [83.75–92.52%]	98.88% [97.89–99.77%]
Within other	11	95.87% [85.84–100.00%]	99.20% [96.02–100.00%]
Belgian vs. other	18	84.41% [83.70–86.00%]	97.42 [96.25–98.13%]
	Nova virus M segment		
Within Belgium	7	85.94% [83.35–90.87%]	97.65% [96.36–99.20%]
Within other	1	N.A.	N.A.
Belgian vs. other	8	85.28% [83.96–85.99%]	97.61% [97.16–97.87%]
	Nova virus L segment		
Within Belgium	7	86.63% [85.50–87.98%]	97.37% [96.34–98.14%]
Within other	2	85.98% [N.A.]	96.38% [N.A.]
Belgian vs. other	9	85.87% [85.42–86.48%]	96.60% [95.96–97.12%]

N.A, not applicable.

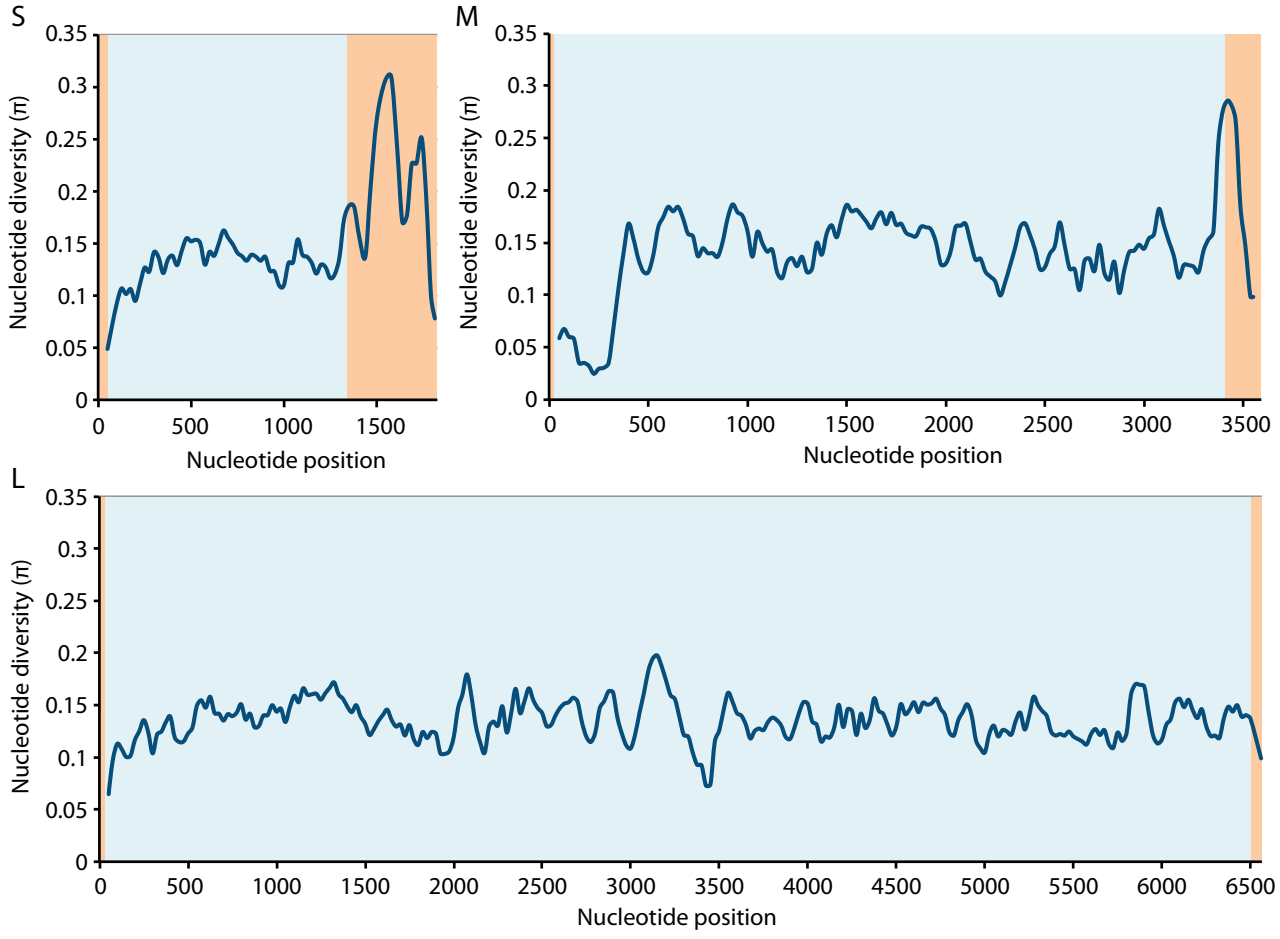

Fig. 2 Interstrain nucleotide diversity (π) plots of complete genomes of Nova virus S, M and L segments. The π values were calculated using a sliding window approach in DNASP with a window size of 100 and a step size of 25. The coding region is marked in blue, while 3′ and 5′ noncoding regions are marked in orange.

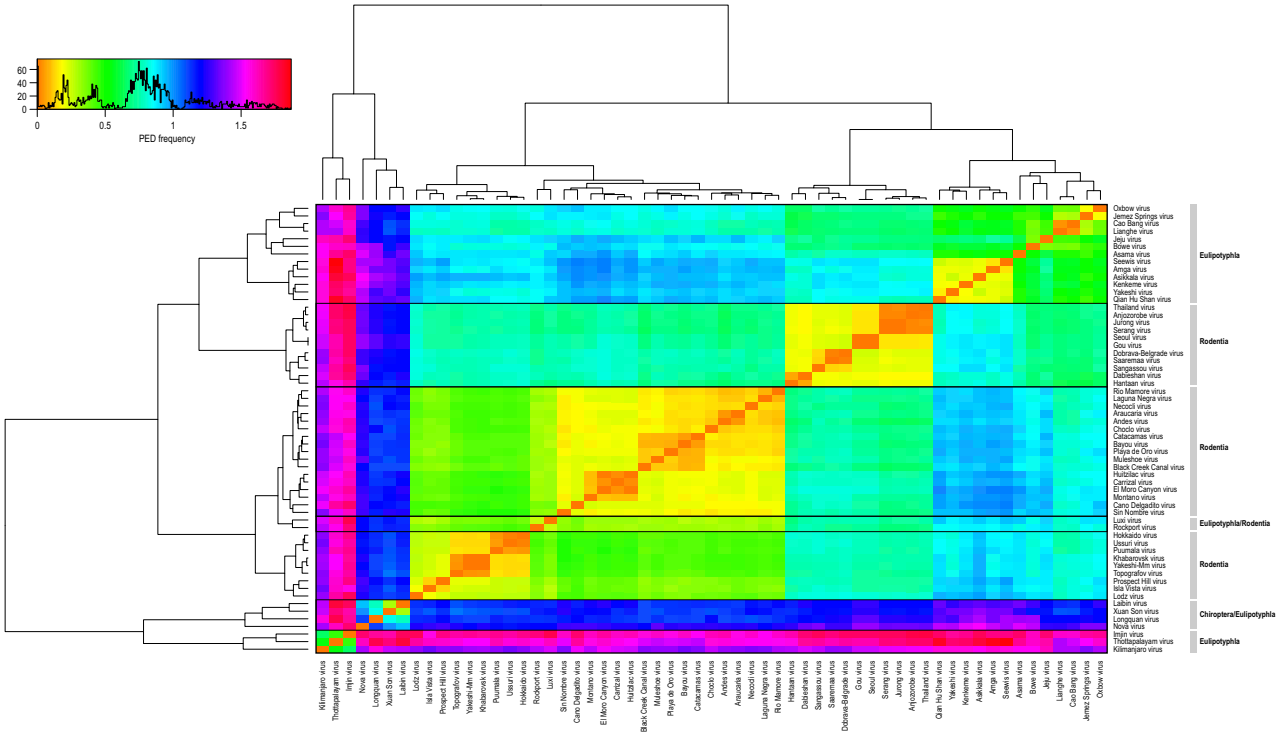

Fig. 3 Heatmap displaying hantavirus genetic divergence. Interstrain nucleotide diversity was calculated using pairwise evolutionary distances (PED) based upon amino acid sequences of the hantavirus nucleocapsid genes. A PED frequency plot shows the used colour scale with corresponding PED values.

divergent group consisted of bat-borne hantaviruses (i.e. Longquan virus, Xuan Son virus and Laibin virus) and Nova virus. Within this group, Nova virus was shown to be the most distantly related hantavirus. Lower intragroup divergence was seen for rodent-borne hantaviruses in contrast to bat-, shrew- and mole-borne hantaviruses. These results indicate that Nova virus is a highly divergent virus and has acquired characteristics, unique among hantaviruses.

The Nova virus nucleocapsid gene was subjected to strong purifying selection

Hantaviruses are thought to have a strict relationship with their natural hosts, putting strong evolutionary constraints on genome variation (Sironen *et al.* 2001; Daugherty & Malik 2012; Castel *et al.* 2014). To detect molecular signatures of selection, we performed a comprehensive selection analysis of the complete coding sequence of the Nova virus S segment, encoding the nucleocapsid gene. For this purpose, the complete S segment of 105 Belgian Nova virus strains, covering the entire sampling region, was sequenced. Of these, five samples identical to other Nova viruses at the nucleotide level and two additional samples with identical coding regions were excluded from further analyses. For the 98 samples included in selection pressure testing, the presence of possible recombination events was excluded by RDP4 (Martin *et al.* 2015). To detect selective pressures acting on the nucleocapsid gene, five different methods (i.e. SLAC, FEL, IFEL, FUBAR and MEME) were employed. The overall dN/dS ratio of the nucleocapsid data set, calculated using SLAC, was 0.006, indicating that Nova virus is subjected to strong purifying selection (Table 2). Negative selection pressures acting on the nucleocapsid gene were detected by SLAC, FEL, IFEL and FUBAR. Depending on the used method, between 320 and 408 of a total of 428 sites were estimated to be negatively selected. This marks between 74.76% and 95.32% of tested codon positions.

When applying SLAC, FEL, IFEL and FUBAR, no positively selected sites could be detected in the nucleocapsid gene. Although we were not able to find evidence for adaptive evolution, applying the aforementioned methods, we screened for positively selected sites under episodic directional selection using MEME. Here, we detected no sites under episodic diversifying selection corresponding *P*-values below 0.05. While values were slightly above the statistical significance level ($P < 0.05$), two sites under possible episodic diversifying selection were detected at codon

Table 2 Selection analysis based on 98 unique coding sequences of the nucleocapsid gene of Nova virus. SLAC, FEL, IFEL, MEME and FUBAR, implemented in the DATAMONKEY server, were used to determine sites under positive, negative or diversifying selection

Method	Cut-off value	Positively selected sites	Sites under episodic diversifying selection	Negatively selected sites	Mean ω
SLAC	$P \leq 0.05$	0/428	N.A.	320/428 (74.76%)	0.006
FEL	$P \leq 0.05$	0/428	N.A.	348/428 (81.31%)	N.A.
IFEL	$P \leq 0.05$	0/428	N.A.	332/428 (77.57%)	N.A.
MEME	$P \leq 0.1$	N.A.	190 ($P = 0.05$), 233 ($P = 0.05$)	N.A.	N.A.
FUBAR	$P \leq 0.9$	0/428	N.A.	408/428 (95.32%)	N.A.

N.A. Not applicable, ω = dN/dS.

positions 190 ($P = 0.05$) and 233 ($P = 0.05$). In our data set, an alanine or proline can be found at position 190, whereas at site 233, a serine or threonine is expressed. Although crystal structures of Nova virus have not been determined, corresponding positions of Hantaan virus flank putative binding regions, whereas the region containing amino acid 190 is thought to be required for a functional RNA binding domain (Xu *et al.* 2002; Olal & Daumke 2016). Although Nova virus exhibits relatively high nucleotide diversity, the nucleocapsid protein has been subjected to strong purifying selection resulting in limited amino acid changes.

Nova virus has efficiently spread in the Belgian mole population

To obtain a broader overview of hantavirus evolution, a Bayesian phylogenetic analysis of the partial nucleotide sequence of the S segment, corresponding to the highest amount of Nova virus sequences available in NCBI GenBank, was performed. The phylogenetic tree shows monophyletic clades, referring to families, subfamilies or orders of their natural hosts (i.e. Muridae, Soricidae, Sigmodontinae, Neotominae, Arvicolinae, Chiroptera; Fig. 4A). Occasionally, cross-species transmission events can be noted. This is especially clear with mole-borne hantaviruses that endured multiple host switches during evolution. Nova viruses form a distant group, most closely related to bat-borne hantaviruses. These results indicate that Nova virus is a highly divergent virus with a complex evolutionary history likely involving cross-species transmission. Bayesian phylogenetic analysis of all available Nova virus strains showed distinct clades according to the geographical location (Fig. 4B). However, Belgian strains form two separate lineages. Nova virus strains corresponding to the west of Belgium are more closely related to strains originating from France, Poland and Hungary, while the east of Belgium represents a separate clade. The Belgian Nova virus strains were further used in a more in-depth phylogenetic reconstruction.

To reconstruct the spatial diffusion of Nova virus in the Belgian mole population, a continuous Bayesian phylogenetic analysis was performed. The sequences are 1286 nt long and span the entire Nova virus nucleocapsid gene. Subsequently, the maximum clade credibility (MCC) tree was mapped according to known and estimated node locations (Fig. 5). Phylogenetic analysis allowed the identification of two distinct and well-supported clades, corresponding to the east and west of Belgium.

Nova virus dispersal was correlated with the geographical location of external and internal nodes. Because the range of sampling dates in our data set was rather limited, we set a fixed estimated substitution rate. The time to the most recent common ancestor (TMRCA), based on a fixed evolutionary rate calibration of 10^{-3} substitutions/site/year, was estimated around the year 1760 (95% highest posterior density interval: 1725–1793). Our phylogenetic analyses revealed that Nova virus has spread across Belgium forming two separate lineages.

The main waterways did not affect Nova virus dispersal velocity in Belgium

To estimate the impact of natural barriers on the dispersal of Nova virus in the mole population, the analytical framework for comparing viral phylogenies with environmental landscape data, developed by Dellicour *et al.* (2016b), was used. The spatio-temporal information of 100 trees, sampled from the posterior distribution of the preceding Bayesian phylogenetic reconstruction of Nova virus in Belgium, was extracted. Subsequently, the association between the main waterways, acting as an environmental resistance factor, and Nova virus dispersal velocity was analysed. For this purpose, a raster containing the main waterways with a grid resolution of 0.5 arcmin and 1 arcmin (corresponding to ~1 km square and ~2 km square, respectively) was employed. For each resistance grid, two different path models, that is a least-cost and a random walk path model, were used in order to compute the environmental weight

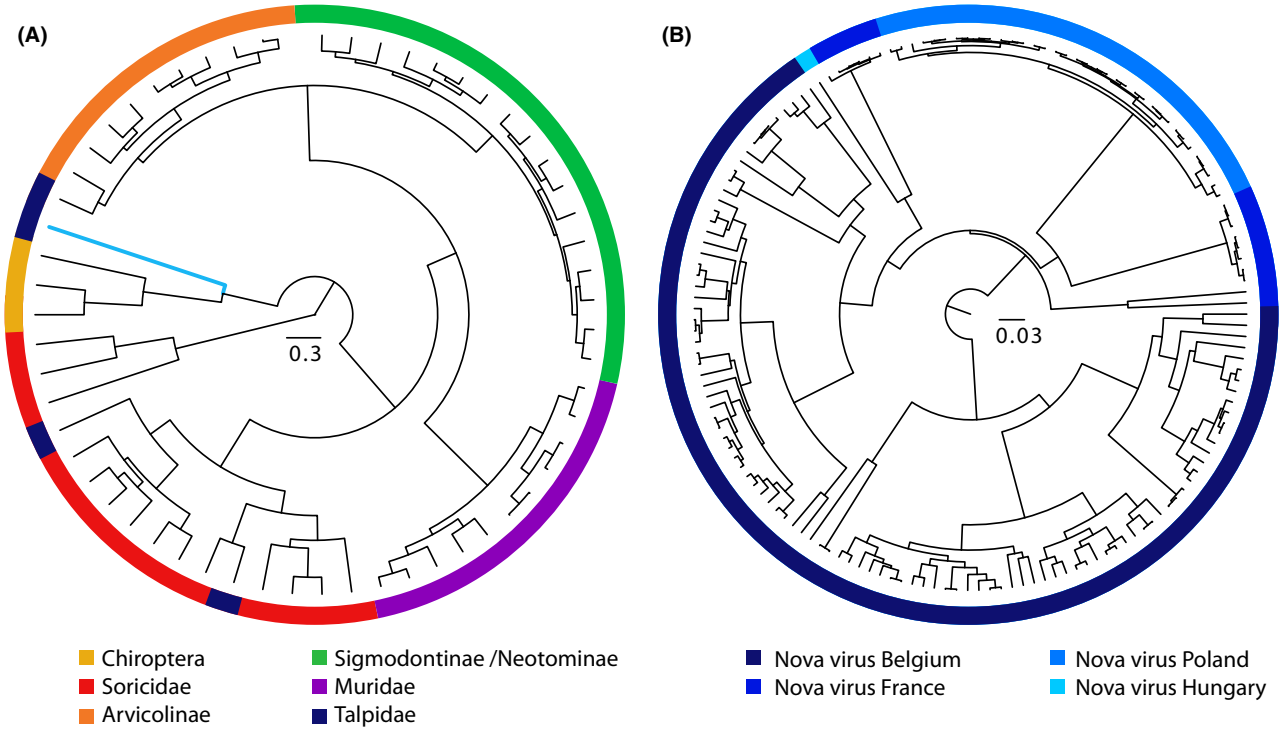

Fig. 4 (A) Bayesian phylogenetic tree of hantavirus partial S segment (447nt). Outer ring is coloured according to the host; Chiroptera in yellow, Soricidae in red, Arvicolinae in orange, Sigmodontinae and Neotominae in green, Muridae in purple and Talpidae in blue. The Nova virus branch is marked in blue. (B) Bayesian phylogenetic tree of the partial S segment (447 nt) of Nova virus. Within the family Talpidae, Nova virus strains originating from different countries (i.e. Belgium, France, Poland and Hungary) are represented in different shades of blue.

assigned to each phylogeny branch. To ensure that a sufficiently conservative approach was employed, a range of *k*-values, used to define the resistance value for the waterways, was chosen. For each combination of grid resolution, *k*-value and path model (Table 3), we estimated the correlation statistic = $(R^2_{\text{waterways}} - R^2_{\text{null}})$, where $R^2_{\text{waterways}}$ is the R^2 obtained when branch durations are regressed against weights computed on a main waterways raster and R^2_{null} is the R^2 obtained when branch durations are regressed against weights computed on a 'null' raster, that is a raster with uniform cell values equal to one (see the Material and Methods section for further details). Only 0 to 2 per cents of *D* statistics were both positive and significant (Table 3). This is well below what would have been expected if the main waterways had an impact on Nova virus dispersal velocity. We can thus conclude that the main waterways in Belgium did not represent a significant resistance factor slowing down the dispersal velocity of Nova virus in the Belgium mole population.

Discussion

The detection of shrew-, bat- and mole-borne hantaviruses illustrates the worldwide spread and high diversity of these viruses, infecting a vast range of small mammals. Nevertheless, nearly a decade after the discovery of a second non-rodent-borne hantavirus, the current limitations to our knowledge concerning evolution, epidemiology and pathogenicity of these hantaviruses are noteworthy (Klempa *et al.* 2007). Although efforts have been made, the paucity of available virus isolates and complete genome sequence data has hampered important aspects of hantavirus research (Shin *et al.* 2012; Ling *et al.* 2015; Gu *et al.* 2016). Here we present an in-depth analysis of the spatial distribution, diversity and evolution of Nova virus, an insectivore-borne hantavirus.

Human hantavirus infection occurs indirectly through aerosols of infectious particles present in the environment. Thus, transmission efficacy of hantavirus is directly dependent upon the shedding behaviour of its natural host. Whereas for insectivore-borne hantaviruses no in vivo studies have been performed, persistence and excretion of rodent-borne hantaviruses have been studied under experimental and natural conditions (Botten *et al.* 2000, 2002; Hardestam *et al.* 2008; Voutilainen *et al.* 2015). While hantavirus shedding usually declines after an initial peak phase, hantavirus RNA is still detected in tissues until at least 7 months

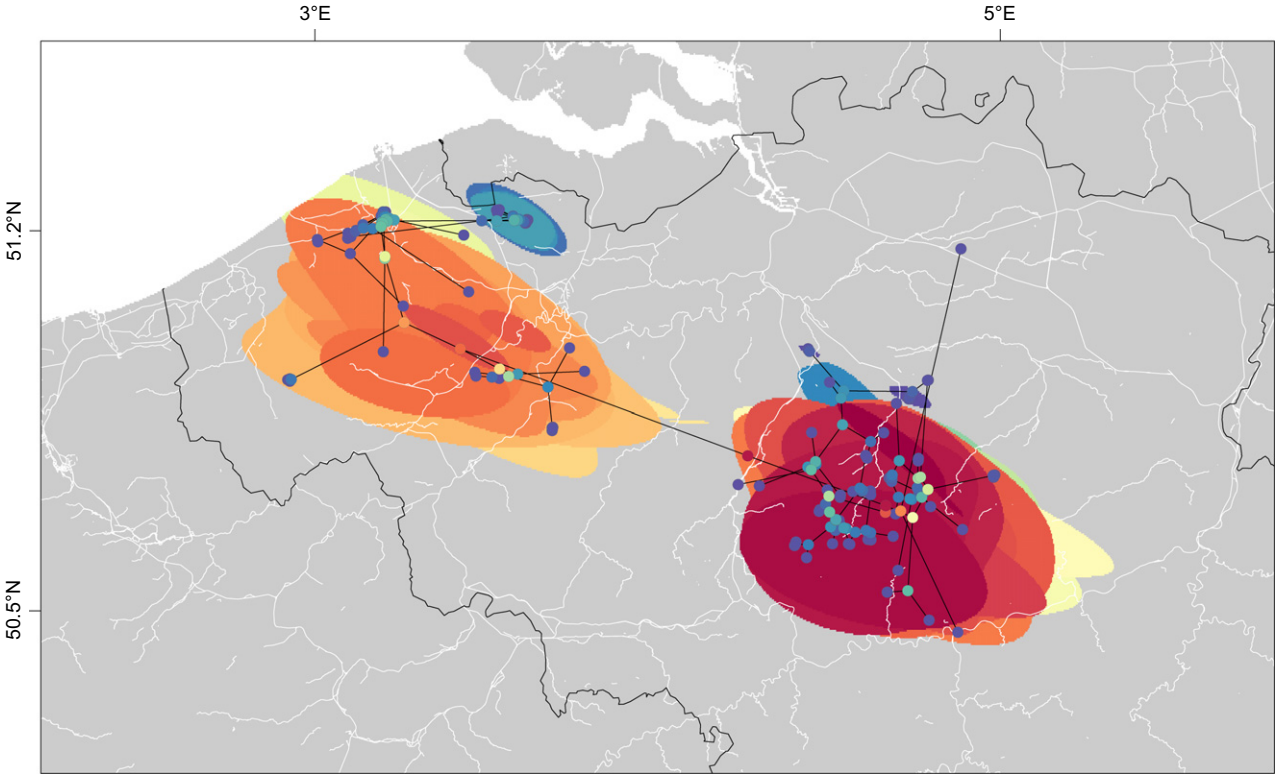

Fig. 5 Reconstructed spatiotemporal diffusion of Nova virus in the Belgian moles population: mapped MCC tree and 75% HPD regions based on 100 trees regularly sampled from the post-burn-in posterior distribution. To discard the large uncertainty related to the position of the most ancestral node connecting the two distinct clades, this node was removed of the sampled trees for the HPD regions estimation (see Figure S1, Supporting information for the map generated without removing the most ancestral node). The 75% HPD regions were computed for successive time layers and then superimposed using a specific colour scale ranging from red (MRCA) to blue (sampling time). The mapped tree is the MCC tree, and its nodes are coloured according to the same colour scale as for the 75% HPD layer. White and black lines correspond to the main waterways ('rivers' and 'canals' as defined by the OPENSTREETMAP database) and to the national Belgian boundaries, respectively.

after infection (Botten *et al.* 2002, 2003). Rodent-borne as well as insectivore-borne hantaviruses are known to have a broad tissue distribution in their reservoir host (Botten *et al.* 2000; Gu *et al.* 2014b; Witkowski *et al.* 2016). Here, we report the detection of Nova virus in 53.2% of all tested mole kidney samples. Hantavirus positivity rates can vary significantly, depending on the host species. Furthermore, within a hantavirus species, seasonal and annual variation in positivity rates are also observed, contingent on multiple factors (e.g. population turnover, population density, host contact rates) (Voutilainen *et al.* 2016). Although moles have a considerably longer lifespan than most rodents, this alone cannot explain the high positivity rate observed in the mole population. In bats, of which many species have a lifespan exceeding 25 years, relatively low hantavirus positivity rates have been reported (Calisher *et al.* 2006; Sumibcay *et al.* 2012; Weiss *et al.* 2012). Although contact with mole excreta will be limited until their capture, we could speculate that a considerable human infection risk could be present, considering the high Nova virus positivity rate. For this purpose, screening of professional mole catchers, gardeners, etc. for the presence of Nova virus IgG antibodies would be a valuable step in the overdue study of the pathogenic potential of this virus.

Nova virus is a highly divergent hantavirus, forming a separate phylogenetic lineage, most closely related to bat-borne hantaviruses. Within the Nova virus species, high nucleotide diversity with circa 29% polymorphic sites at the S, M and L segments is present. The S segment, similar to other hantaviruses, contains a large 3′ noncoding region (circa 500 nt) that has limited sequence conservation. The 3′ NCR of hantavirus S segment mRNA is predicted to be involved in cap-dependent translation initiation of viral mRNAs, replacing polyA tail functions (Vera-Otarola *et al.* 2010). Furthermore, mutations in 3′ and 5′ NCRs are associated with Puumala virus-host adaptation (Lundkvist *et al.* 1997; Nemirov *et al.* 2003). Although the 3′ NCR of Belgian

Table 3 Results of the randomization tests performed to assess the level of significance of the correlation between phylogenetically informed dispersal durations and environmental weights computed on the main waterways raster. Randomization tests were applied to a set of 100 sampled trees, and 100 randomization steps were performed to assess the significance of the *D* statistic estimated for each tree. For each combination of raster grid resolution, *k*-value and path taken model, we report the percentage of sampled trees for which $P < 0.05$. '*k*' refers to the parameter used to define the resistance value of main waterways acting as barriers

Resolution	*k*	Path taken model	% of significant positive *D* values
0.008 × 0.008 (0.5 × 0.5 arcmin)	10	Least-cost path	2
		Random walk path	1
	100	Least-cost path	1
		Random walk path	0
	1000	Least-cost path	0
		Random walk path	0
0.016 × 0.016 (1 × 1 arcmin)	10	Least-cost path	0
		Random walk path	0
	100	Least-cost path	0
		Random walk path	0
	1000	Least-cost path	0
		Random walk path	0

Nova virus strains is highly variable, two relatively conserved domains are present around nucleotide 1400 and nucleotide 1650. Whether these regions are involved in RNA binding of host or viral proteins remains to be further investigated. The M segment contains a conserved region of circa 400 nt that corresponds to the signal peptide and the N-terminal region of the Gn protein. The presence of a less variable region at the N-terminal domain of Gn is remarkable, considering this region corresponds to the ectodomain of Gn, that is projected at the virion surface and is the target of neutralizing antibodies (Arikawa *et al.* 1989; Maes *et al.* 2004). Furthermore, the Gn ectodomain of Nova virus appears to be divergent in comparison with other hantaviruses.

Most hantaviruses have four predicted N-linked glycosylation sites in the Gn protein and one additional N-glycosylation site at the Gc protein in common. These sites, corresponding to N134, N235, N347, N399 and N928 sites of Hantaan virus, were experimentally confirmed to play a role in protein folding and intracellular trafficking (Shi & Elliott 2004). In addition to these corresponding sites, Nova virus has another predicted glycosylation site at N539, whereof the analogue Hantaan virus site was demonstrated in vitro not to be glycosylated during maturation (Shi & Elliott 2004). Finally, for all Nova virus strains, another N-linked glycosylation site, at position N101 in the conserved N-terminal Gn region, was predicted. This region was demonstrated to be significantly divergent from other hantaviruses. The meaning of this finding for Nova virus glycoprotein structure and folding or intracellular trafficking remains to be elucidated.

While hantaviruses have a very strict relationship with their host, assuming to have co-evolved over millions of years, hantavirus evolution has also been shaped by multiple host-switching events (Yanagihara *et al.* 2014). Virus-host co-evolution implements a fine balance between adaptation and counter-adaptation during the host-virus arms race (Daugherty & Malik 2012). Selection analysis of the Nova virus nucleocapsid gene demonstrated that while Nova virus is highly variable at the nucleotide level, it has been subjected to strong purifying selection, illustrated by a dN/dS ratio of 0.006. While no positive selection was detected in our data set using SLAC, FEL, IFEL and FUBAR, we did see indications for diversifying positive selection. It has been hypothesized that adaptive evolution can be primarily episodic, with purifying or neutral selection on other branches masking transient selection on a subset of the lineage (Murrell *et al.* 2012). We can conclude that purifying selection is the dominant pressure driving virus-host co-evolution. A likely scenario would be that after a host-switching event, hantaviruses could undergo a short burst of positive selection, while adapting to the new host. Other studies with rodent-borne hantaviruses previously reported substitution rates within the range 10^{-2} to 10^{-4} subs/site/year, as expected for RNA viruses (Ramsden *et al.* 2008; Weber de Melo *et al.* 2015). While hantaviruses were shown to be subjected to genetic drift and reassortment at a local scale, part of the variation was transient and balancing selection, limiting amino acid changes, was also noticed with other hantaviruses (Razzauti *et al.* 2013; Castel *et al.* 2014; Weber de Melo *et al.* 2015).

In this study, we conducted a phylogenetic reconstruction of Nova virus dispersal in the European mole population in Belgium. Two distinct and well-supported clades, corresponding to the east and west of

Belgium, were defined. Mitochondrial DNA analysis of European mole samples previously found evidence for the presence of three lineages of *Talpa europaea*. Of these, two were restricted to Italy and Spain and a third was demonstrated to be widespread across Europe. Because all European moles that tested Nova virus positive appear to belong to the same European lineage, the presence of two distinct Nova virus lineages can thus not be explained through mitochondrial DNA data (Feuda *et al.* 2015). We concluded that Nova virus has a widespread distribution in Belgium. Previous studies reported different theories concerning hantavirus evolution. The initial hypothesis of the co-evolution of rodent-borne hantaviruses with their hosts over millions of years has been well described (Plyusnin *et al.* 1996). This theory was challenged by Ramsden *et al.* (2009), giving the process of preferential host-switching and local adaptation as preferred explanation for the apparent similarities between phylogenies of hantaviruses and their hosts and hinting to a far more recent hantavirus origin. Recently, the most widely accepted theory is one of virus-host codivergence and virus host-switching both playing an important role in hantavirus evolution (Guo *et al.* 2013). While our TMRCA calculations go back as far as 1760, the strong purifying selection, detected for the nucleocapsid protein, and possible saturation of nucleotide sites make molecular dating of these viruses impossible (Wertheim & Kosakovsky Pond 2011).

Phylogenetic reconstruction demonstrated that Nova virus has spread very efficiently in the European mole population in Belgium. Furthermore, we found that the main waterways in Belgium did not slow down Nova virus dispersal. Such landscape heterogeneity does not limit Nova virus, which can rapidly infect new host animals. Whether this is explained by the ability of moles to swim over short distances or the possibility of ancient dispersal during mole colonization, which is not picked up in our analysis, remains a question.

Nova virus is a mole-borne hantavirus of which the pathogenicity is not yet known. This study provides a comprehensive analysis of Nova virus genetic diversity, evolution and spatial spread. Although Nova virus is highly divergent, conserved regions were present in the Nova virus genome. Furthermore, Nova virus underwent strong purifying selection, leading to a conserved nucleocapsid protein. Nova virus has a high positivity rate and has efficiently spread in the mole population. A next step would be to study the pathogenic potential of this mole-borne virus. Here we provide more information about ecological factors driving hantavirus diffusion. Furthermore, we provide a source of sequence information that is of value for subsequent functional experiments.

Acknowledgements

This work was supported by a BOF-CREA grant from the KU Leuven (3M110189). We are grateful to Sam Vanhille, Ghiseleen Pauwels, Isabelle Hoflack, Guy Blondiau and Pascal Lethem, Dirk Huybrechts and Femke Bleys for their knowledge and/or technical assistance with fieldwork.

References

Amori G, Hutterer R, Mitsain G *et al.* (2008) *Talpa europaea*. *The IUCN Red List of Threatened Species 2008*: e.T41481A10462965. http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T41481A10462965.en

Arai S, Ohdachi SD, Asakawa M *et al.* (2008) Molecular phylogeny of a newfound hantavirus in the Japanese shrew mole (*Urotrichus talpoides*). *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 16296–16301.

Arikawa J, Schmaljohn AL, Dalrymple JM, Schmaljohn CS (1989) Characterization of Hantaan virus envelope glycoprotein antigenic determinants defined by monoclonal antibodies. *Journal of General Virology*, **70**(Pt 3), 615–624.

Bennett SN, Gu SH, Kang HJ, Arai S, Yanagihara R (2014) Reconstructing the evolutionary origins and phylogeography of hantaviruses. *Trends in Microbiology*, **22**, 473–482.

Bielejec F, Rambaut A, Suchard MA, Lemey P (2011) SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics. *Bioinformatics*, **27**, 2910–2912.

Botten J, Mirowsky K, Kusewitt D *et al.* (2000) Experimental infection model for Sin Nombre hantavirus in the deer mouse (*Peromyscus maniculatus*). *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 10578–10583.

Botten J, Mirowsky K, Ye C *et al.* (2002) Shedding and intracage transmission of Sin Nombre hantavirus in the deer mouse (*Peromyscus maniculatus*) model. *Journal of Virology*, **76**, 7587–7594.

Botten J, Mirowsky K, Kusewitt D *et al.* (2003) Persistent Sin Nombre virus infection in the deer mouse (*Peromyscus maniculatus*) model: sites of replication and strand-specific expression. *Journal of Virology*, **77**, 1540–1550.

Brocato RL, Hammerbeck CD, Bell TM *et al.* (2014) A lethal disease model for hantavirus pulmonary syndrome in immunosuppressed Syrian hamsters infected with Sin Nombre virus. *Journal of Virology*, **88**, 811–819.

Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T (2006) Bats: important reservoir hosts of emerging viruses. *Clinical Microbiology Reviews*, **19**, 531–545.

Carey DE, Reuben R, Panicker KN, Shope RE, Myers RM (1971) Thottapalayam virus: a presumptive arbovirus isolated from a shrew in India. *Indian Journal of Medical Research*, **59**, 1758–1760.

Castel G, Razzauti M, Jousselin E, Kergoat GJ, Cosson JF (2014) Changes in diversification patterns and signatures of selection during the evolution of murinae-associated hantaviruses. *Viruses*, **6**, 1112–1134.

Cleaveland S, Laurenson MK, Taylor LH (2001) Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philosophical*

Transactions of the Royal Society of London. Series B, Biological sciences, **356**, 991–999.

Clement J, Maes P, Lagrou K, Van Ranst M, Lameire N (2012) A unifying hypothesis and a single name for a complex globally emerging infection: hantavirus disease. *European Journal of Clinical Microbiology and Infectious Diseases*, **31**, 1–5.

Daugherty MD, Malik HS (2012) Rules of engagement: molecular insights from host-virus arms races. *Annual Review of Genetics*, **46**, 677–700.

Dellicour S, Rose R, Faria NR, Lemey P, Pybus OG (2016a) SERAPHIM: studying environmental rasters and phylogenetically informed movements. *Bioinformatics*, **32**, 3204–3206.

Dellicour S, Rose R, Pybus OG (2016b) Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. *BMC Bioinformatics*, **17**, 82.

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, **26**, 2455–2457.

Dijkstra EW (1959) A note on two problems in connexion with graphs. *Numerische Mathematik*, **1**, 269–271.

Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, **29**, 1969–1973.

Feuda R, Bannikova AA, Zemlemerova ED *et al.* (2015) Tracing the evolutionary history of the mole, *Talpa europaea*, through mitochondrial DNA phylogeography and species distribution modelling. *Biological Journal of the Linnean Society*, **114**, 495–512.

Gill MS, Lemey P, Faria NR *et al.* (2013) Improving Bayesian population dynamics inference: a coalescent-based model for multiple loci. *Molecular Biology and Evolution*, **30**, 713–724.

Gizzi M, Delaere B, Weynand B *et al.* (2013) Another case of "European hantavirus pulmonary syndrome" with severe lung, prior to kidney, involvement, and diagnosed by viral inclusions in lung macrophages. *European Journal of Clinical Microbiology and Infectious Diseases*, **32**, 1341–1345.

Gu SH, Dormion J, Hugot JP, Yanagihara R (2014a) High prevalence of Nova hantavirus infection in the European mole (*Talpa europaea*) in France. *Epidemiology and Infection*, **142**, 1167–1171.

Gu SH, Hejduk J, Markowski J *et al.* (2014b) Co-circulation of soricid- and talpid-borne hantaviruses in Poland. *Infection, Genetics and Evolution*, **28**, 296–303.

Gu SH, Kumar M, Sikorska B *et al.* (2016) Isolation and partial characterization of a highly divergent lineage of hantavirus from the European mole (*Talpa europaea*). *Scientific Reports*, **6**, 21119.

Guo WP, Lin XD, Wang W *et al.* (2013) Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents. *PLoS Pathogens*, **9**, e1003159.

Hardestam J, Karlsson M, Falk KI *et al.* (2008) Puumala hantavirus excretion kinetics in bank voles (*Myodes glareolus*). *Emerging Infectious Diseases*, **14**, 1209–1215.

Hughes AL, Friedman R (2000) Evolutionary diversification of protein-coding genes of hantaviruses. *Molecular Biology and Evolution*, **17**, 1558–1568.

Jones KE, Patel NG, Levy MA *et al.* (2008) Global trends in emerging infectious diseases. *Nature*, **451**, 990–993.

Kang HJ, Bennett SN, Dizney L *et al.* (2009a) Host switch during evolution of a genetically distinct hantavirus in the American shrew mole (*Neurotrichus gibbsii*). *Virology*, **388**, 8–14.

Kang HJ, Bennett SN, Sumibcay L *et al.* (2009b) Evolutionary insights from a genetically divergent hantavirus harbored by the European common mole (*Talpa europaea*). *PLoS ONE*, **4**, e6149.

Kang HJ, Bennett SN, Hope AG, Cook JA, Yanagihara R (2011) Shared ancestry between a newfound mole-borne hantavirus and hantaviruses harbored by cricetid rodents. *Journal of Virology*, **85**, 7496–7503.

Kang HJ, Gu SH, Cook JA, Yanagihara R (2016) Dahonggou Creek virus, a divergent lineage of hantavirus harbored by the long-tailed mole (*Scaptonyx fusicaudus*). *Tropical Medicine and Health*, **44**, 16.

Klempa B, Fichet-Calvet E, Lecompte E *et al.* (2006) Hantavirus in African wood mouse, Guinea. *Emerging Infectious Diseases*, **12**, 838–840.

Klempa B, Fichet-Calvet E, Lecompte E *et al.* (2007) Novel hantavirus sequences in Shrew, Guinea. *Emerging Infectious Diseases*, **13**, 520–522.

Kosakovsky Pond SL, Frost SD (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, **22**, 1208–1222.

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, **33**, 1870–1874.

Lemey P, Rambaut A, Welch JJ, Suchard MA (2010) Phylogeography takes a relaxed random walk in continuous space and time. *Molecular Biology and Evolution*, **27**, 1877–1885.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Lin XD, Zhou RH, Fan FN *et al.* (2014) Biodiversity and evolution of Imjin virus and Thottapalayam virus in Crocidurinae shrews in Zhejiang Province, China. *Virus Research*, **189**, 114–120.

Ling J, Vaheri A, Hepojoki S *et al.* (2015) Serological survey of Seewis virus antibodies in patients suspected for hantavirus infection in Finland; a cross-reaction between Puumala virus antiserum with Seewis virus N protein? *Journal of General Virology*, **96**, 1664–1675.

Lundkvist A, Cheng Y, Sjolander KB *et al.* (1997) Cell culture adaptation of Puumala hantavirus changes the infectivity for its natural reservoir, *Clethrionomys glareolus*, and leads to accumulation of mutants with altered genomic RNA S segment. *Journal of Virology*, **71**, 9515–9523.

Maes P, Clement J, Gavrilovskaya I, Van Ranst M (2004) Hantaviruses: immunology, treatment, and prevention. *Viral Immunology*, **17**, 481–497.

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, **1**, vev003.

McRae BH (2006) Isolation by resistance. *Evolution*, **60**, 1551–1561.

Mcrae BH, Dickson BG, Keitt TH, Shah VB (2008) Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, **89**, 2712–2724.

Murrell B, Wertheim JO, Moola S *et al.* (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, **8**, e1002764.

Murrell B, Moola S, Mabona A *et al.* (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, **30**, 1196–1205.

Nemirov K, Lundkvist A, Vaheri A, Plyusnin A (2003) Adaptation of Puumala hantavirus to cell culture is associated with point mutations in the coding region of the L segment and in the noncoding regions of the S segment. *Journal of Virology*, **77**, 8793–8800.

Olal D, Daumke O (2016) Structure of the hantavirus nucleoprotein provides insights into the mechanism of RNA encapsidation. *Cell Reports*, **14**, 2092–2099.

Peeraer K, Couck I, Debrock S *et al.* (2015) Frozen-thawed embryo transfer in a natural or mildly hormonally stimulated cycle in women with regular ovulatory cycles: a RCT. *Human Reproduction*, **30**, 2552–2562.

Peters CJ, Khan AS (2002) Hantavirus pulmonary syndrome: the new American hemorrhagic fever. *Clinical Infectious Diseases*, **34**, 1224–1231.

Plyusnin A, Vapalahti O, Vaheri A (1996) Hantaviruses: genome structure, expression and evolution. *Journal of General Virology*, **77**(Pt 11), 2677–2687.

Plyusnin A, Beaty BJ, Elliott RM *et al.* (2011) Bunyaviridae. In: *Virus Taxonomy: 9th Report of the International Committee on Taxonomy of Viruses* (eds King AM, Lefkowitz E, J. AM , Carstens EB), pp. 722–741. Academic Press, London, UK.

Pond SL, Frost SD, Grossman Z *et al.* (2006) Adaptation to different human populations by HIV-1 revealed by codon-based analyses. *PLoS Computational Biology*, **2**, e62.

Pybus OG, Suchard MA, Lemey P *et al.* (2012) Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 15066–15071.

Rambaut A, Lam TT, de Carvalho L, Pybus OG (2016) Exploring the temporal structure of heterochronous sequences using TempEst. *Virus Evolution*, **2**, vew007.

Ramsden C, Melo FL, Figueiredo LM *et al.* (2008) High rates of molecular evolution in hantaviruses. *Molecular Biology and Evolution*, **25**, 1488–1492.

Ramsden C, Holmes EC, Charleston MA (2009) Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. *Molecular Biology and Evolution*, **26**, 143–153.

Razzauti M, Plyusnina A, Henttonen H, Plyusnin A (2013) Microevolution of Puumala hantavirus during a complete population cycle of its host, the bank vole (*Myodes glareolus*). *PLoS ONE*, **8**, e64447.

Schlegel M, Radosa L, Rosenfeld UM *et al.* (2012) Broad geographical distribution and high genetic diversity of shrew-borne Seewis hantavirus in Central Europe. *Virus Genes*, **45**, 48–55.

Shapiro B, Rambaut A, Drummond AJ (2006) Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. *Molecular Biology and Evolution*, **23**, 7–9.

Shi X, Elliott RM (2004) Analysis of N-linked glycosylation of hantaan virus glycoproteins and the role of oligosaccharide side chains in protein folding and intracellular trafficking. *Journal of Virology*, **78**, 5414–5422.

Shin OS, Yanagihara R, Song JW (2012) Distinct innate immune responses in human macrophages and endothelial cells infected with shrew-borne hantaviruses. *Virology*, **434**, 43–49.

Sironen T, Vaheri A, Plyusnin A (2001) Molecular evolution of Puumala hantavirus. *Journal of Virology*, **75**, 11803–11810.

Steentoft C, Vakhrushev SY, Joshi HJ *et al.* (2013) Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *EMBO Journal*, **32**, 1478–1488.

Sumibcay L, Kadjo B, Gu SH *et al.* (2012) Divergent lineage of a novel hantavirus in the banana pipistrelle (*Neoromicia nanus*) in Cote d'Ivoire. *Virology Journal*, **9**, 34.

Taylor LH, Latham SM, Woolhouse ME (2001) Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society of London. Series B, Biological sciences*, **356**, 983–989.

Vaheri A, Strandin T, Hepojoki J *et al.* (2013) Uncovering the mysteries of hantavirus infections. *Nature Reviews Microbiology*, **11**, 539–550.

Vera-Otarola J, Soto-Rifo R, Ricci EP *et al.* (2010) The 3′ untranslated region of the Andes hantavirus small mRNA functionally replaces the poly(A) tail and stimulates cap-dependent translation initiation from the viral mRNA. *Journal of Virology*, **84**, 10420–10424.

Voutilainen L, Sironen T, Tonteri E *et al.* (2015) Life-long shedding of Puumala hantavirus in wild bank voles (*Myodes glareolus*). *Journal of General Virology*, **96**, 1238–1247.

Voutilainen L, Kallio ER, Niemimaa J, Vapalahti O, Henttonen H (2016) Temporal dynamics of Puumala hantavirus infection in cyclic populations of bank voles. *Scientific Reports*, **6**, 21323.

Weber de Melo V, Sheikh Ali H, Freise J *et al.* (2015) Spatiotemporal dynamics of Puumala hantavirus associated with its rodent host, *Myodes glareolus*. *Evolutionary Applications*, **8**, 545–559.

Weiss S, Witkowski PT, Auste B *et al.* (2012) Hantavirus in bat, Sierra Leone. *Emerging Infectious Diseases*, **18**, 159–161.

Wertheim JO, Kosakovsky Pond SL (2011) Purifying selection can obscure the ancient age of viral lineages. *Molecular Biology and Evolution*, **28**, 3355–3365.

Witkowski PT, Drexler JF, Kallies R *et al.* (2016) Phylogenetic analysis of a newfound bat-borne hantavirus supports a laurasiatherian host association for ancestral mammalian hantaviruses. *Infection, Genetics and Evolution*, **41**, 113–119.

Woolhouse ME, Gowtage-Sequeria S (2005) Host range and emerging and reemerging pathogens. *Emerging Infectious Diseases*, **11**, 1842–1847.

Xu X, Severson W, Villegas N, Schmaljohn CS, Jonsson CB (2002) The RNA binding domain of the hantaan virus N protein maps to a central, conserved region. *Journal of Virology*, **76**, 3301–3308.

Yanagihara R, Gu SH, Arai S, Kang HJ, Song JW (2014) Hantaviruses: rediscovery and new beginnings. *Virus Research*, **187**, 6–14.

Data accessibility

All sequences generated in this study can be found at NCBI GenBank under Accession nos KX512326 to KX512437.

L.L. and P.M. designed the study. L.L., V.V. and P.M. collected samples. L.L., V.V., S.D.C., I.N., I.V. and B.V. performed the experiments. L.L., S.D., P.L. and P.M. analysed the data. L.L., S.D. and P.M. wrote the manuscript. All authors reviewed the manuscript.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Primer sequences used to amplify or sequence the complete S segment of Nova virus. Thermal profiles are available upon request.

Table S2 Predicted O-linked glycosylation sites of Nova virus glycoproteins using the NetOGlyc 4.0.0.13 server (CBS prediction server, DTU).

Fig. S1 Reconstructed spatiotemporal diffusion of Nova virus in the Belgian moles population: mapped MCC tree and 75% HPD regions based on 100 trees regularly sampled from the post burn-in posterior distribution.